



## Web Harvesting Pilot Project

GPO is pleased to announce the release of a white paper on the results of the recently completed Web Harvesting pilot project to capture official Environmental Protection Agency (EPA) publications in scope of GPO's information dissemination programs.



### [Web Harvesting White Paper](#)

The white paper reports on the specific context of the results of the pilot, including a summary of analysis done on the work performed, an assessment of lessons learned, and planned future direction and next steps for further development of the harvesting function to be implemented during Release 2 of GPO's Future Digital System (FDsys), currently scheduled for mid-2008.

As a first step in learning about automated Web publication discovery as well as harvesting technologies and methodologies, GPO contracted with two private companies on this pilot. The collaboration developed rules and instructions that would determine whether EPA content discovered was in scope for GPO's dissemination programs. Three separate crawls were conducted on the sites over a six-month period. Harvester rules and instructions were refined and revised between crawls.